DYNAMIC SIGNALING AND ROUTING

FIELD OF THE INVENTION

The present invention relates generally to communication networks and particularly to methods of controlling real time data transmission over packet based networks.

BACKGROUND OF THE INVENTION

Packet based network infrastructure service providers are naturally interested in maximizing traffic over their network. In order to attract clients, it is required to provide sufficient quality of service (QoS). For real time connections, such as video conferencing and telephone connections, providing sufficient QoS includes providing the packets of the connection within a predetermined time (low delay) from end to end, with a low variance in arrival time (jitter) and having a low packet loss rate. One major factor which affects the provided QoS is the load on the links of the network. Therefore, it has been suggested, for example in US patent publication 2002/0110112 to Jukka Tuomi, and US patent 6,553,515 to Gross et al., the disclosures of which are incorporated herein by reference, that service providers limit the maximal number of connections allowed to pass concurrently over the network and/or prevent establishment of new connections when the network is congested. Naturally, however, it is desired to accept a maximal number of connections in order to maximize revenues.

In order to allow better utilization of the network, routing methods attempt to pass connections on relatively less congested paths of the network. U.S. Patent 4,905,233, the disclosure of which is incorporated herein by reference, describes the selection of a route based on a link metric which is calculated from the current data rate, the link capacity (both in packets per second), and delay over the link.

Japanese patent publication JP2002/359651, the disclosure of which is incorporated herein by reference, describes a network management device that periodically collects network information and accordingly determines which codec is to be used for newly established video conference sessions. Media gateways contact the network management device when a connection is to be established and receive from the device instructions on the codec they are to use.

U.S. patent 6,529,475 to Wan et al., the disclosure of which is incorporated herein by reference, describes a network in which monitors extract from the real time control protocol (RTCP) packets information regarding network congestion. A traffic control mechanism utilizes the extracted information to determine whether the network is congested. In case the

network is congested, the control mechanism instructs gatekeepers of the network to perform actions which reduce congestion. The 6,529,475 patent suggests reducing the congestion either by limiting the number of new calls accepted to pass on the network or by bandwidth reduction of currently active calls. The bandwidth reduction is performed either by instructing the end-point applications to reduce their bandwidth or by altering the content of the RTCP packets provided to the end-point applications so that the applications reduce their bandwidth consumption. The patent suggests reducing the bandwidth for video by adjusting window size, frame rate, video quality and color coding schemes. In voice-only connections, the 6,529,475 patent states that "the bandwidth reduction options are limited. For example, the lowest coding rate for VoIP is 5.3 kbps which, including all the header overhead, has an effective rate of only 16 kbps. There are ways to reduce it further, such as silence compression and header compression."

Some end-points are programmed to adjust their transmission parameters according to the conditions of their connection. Methods of operation for such end-points are described, for example, in U.S. patent publication 2003/0099298 to Rose et al., U.S. patent 6,046,985 to Aldred et al., and U.S. patent 6,356,545 to Vargo et al., the disclosures of which are incorporated herein by reference. The 6,356,545 patent, for example, describes a network in which voice ports transmitting telephone signals over the network adjust the compression/decompression (codec) method according to network conditions. The network conditions are determined based on the quality of voice packets received by the voice ports. In addition to changing the codec, other parameters of the voice ports are adjusted, such as the level of error correction redundancy, the packet size and the packet bundling. U.S. patent 6,356,545 further states that the choice of the codec at the transmitter may be derived from a complex function of choices of packet redundancy, packet size and packet bundling. As an example, the patent suggests simultaneously increasing the packet redundancy and changing the codec.

The optimization criteria of the service provider and the end user may be very different. While the user wants to maximize quality and minimize costs, the service provider wants to maximize revenues by maximizing the number of connections serviced, without causing clients to be dissatisfied.

As discussed above, the handling of the signals of each connection may be adjusted by setting various parameters that affect the connection. In the following discussion these parameters are divided into end-point parameters and network parameters.

The end-point parameters affect the way the end-points of the connection prepare and handle the packets they transmit and receive. The end-point parameters are further divided herein to negotiated parameters and non-negotiated parameters. The values used for the negotiated parameters are selected based on an exchange of signaling packets between the end-points and include, for example, the type of the codec used. Non-negotiated parameters are selected in the art by each end-point separately, without relation to the other end-point, and include, for example, the jitter buffer size.

The network parameters affect the handling of the packets of the connection by the network between the end-points, and include, for example, the route of the packets, the maximal transmission unit (MTU) of the links of the network, the QoS method implemented by the routing units and/or the type of header compression used by the network (if at all). The network parameters may be classified as per-connection parameters or global parameters that affect a plurality of connections.

It is noted that some network parameters are implemented by putting the selected parameter value in the transmitted packets of the connection, so that routers along the path of the packets can operate according to the parameter value. These parameters are considered herein as being network parameters as they affect the way the network handles the packets of the connection.

Some of the network and end-point parameters, such as the route and the codec, affect both the QoS of the specific connection and the total number of connections that can be handled by the network. Other parameters, such as the jitter buffer size and the QoS rating of the connection, affect the connection, but do not affect the total capacity of the network.

SUMMARY OF THE INVENTION

An aspect of some embodiments of the invention relates to joint selection of values of at least one network parameter and at least one end-point parameter, for a real-time packet based connection. The joint selection is optionally performed by the same entity and/or based on the same information or information collected by the same entity. In some embodiments of the invention, the at least one network parameter and the at least one end-point parameter are implemented by different units. The network parameter is optionally implemented by one or more routing units of the network, while the end-point parameter is optionally implemented by one or both of the end-points of the connection. Optionally, the values of the end-point and network parameters are selected before either of the parameter values is implemented. Joint selection of parameter values which are implemented by different units, although requiring

cooperation between different units that set the parameter values, allows for selecting values which better optimize the network utilization.

In some embodiments of the invention, the selection of the network and end-point parameter values is inter-related, i.e., the value selected for a network parameter is taken into account in setting the value for an end-point parameter or *vice versa*. For example, a connection may be passed on a route having a relatively long delay, but with a lower bandwidth utilization, when it is known that the connection has a short jitter buffer which incurs a relatively short delay, so that the total delay on the packets of the connection is below a noticeable delay. In some embodiments of the invention, the end-point parameters of the connection are selected responsive to the route of the packets of the connection. For example, a connection which passes on a route having a short delay, which is very popular, or a route that is loaded, is optionally assigned to use a strong compression which minimizes the bandwidth used by the connection.

Optionally, the parameter value selection is performed based on real time network information collected by a network entity which is not an end-point of the connection. Furthermore, the joint selection of parameter values is optionally performed by a network entity which is not an end-point of the connection. In some embodiments of the invention, the network entity selecting the parameter values performs the selection for a plurality of different connections.

Optionally, joint selection of parameters is performed for each connection at the time the connection is established. In some embodiments of the invention, the real-time network information used for a specific connection is collected without relation to the specific connection, for example before the request to establish the specific connection.

In some embodiments of the invention, the joint selection of parameter values is performed for all connections of the network regardless of the load on the network. Alternatively, the joint selection of parameter values is performed only when the load on the network is relatively high and/or when the network otherwise requires the selection of parameters or is expected to require such selection of parameters. In other cases, the parameter values are not selected jointly.

In some embodiments of the invention, at least one of the network parameters whose value is selected affects the total capacity of the network. Optionally, the value of at least one of the network parameters is not conveyed to the network routing units implementing the parameter through all the packets of the connection which are affected by the parameter value.

For example, the parameter value may be conveyed by dedicated control packets or by one or more leading packets of the connection.

An aspect of some embodiments of the invention relates to selecting non-negotiated end-point parameters for a connection being established over a network, based on real time network information collected by a network entity which is not an end-point of the connection. The non-negotiated end-point parameters include, for example, frame size, voice sampling rate, video resolution, an end-point header compression scheme to be used and/or jitter buffer size. Setting the end-point parameters based on network information allows for selecting parameter values which optimize the network utilization, rather than optimize each connection separately.

Optionally, the real-time network information used in determining parameter values for a specific connection is collected without relation to the specific connection, for example before the request to establish the specific connection. In such cases, the selection does not delay the establishment of the connection. The real time network information optionally comprises quality of service (QoS) conditions of segments and/or routes along the network, such as jitter, packet delay, bandwidth consumption and packet loss rate.

An aspect of some embodiments of the invention relates to joint selection of parameter values for a plurality of different end-point parameters of a real time connection being established over a network, by a network entity other than the end-point. The end-point parameters include, for example, codec type, frame size and jitter buffer size. For video connections, various compression related parameters may be used, such as image resolution and frame rate.

The parameter value selection is optionally performed for all connections established through the network regardless of the load on the network.

An aspect of some embodiments of the invention relates to joint selection of values for one or more end-point parameters of a plurality of different connections passing through a network. Optionally, the joint selection of parameter values is performed for a plurality of connections passing through a single routing unit (e.g., router, switch). In some embodiments of the invention, the parameter value selection for different connections is inter-related, such that the parameter value for one or more of the connections depends on the parameter value for one or more other connections. Optionally, the joint selection includes assigning different values of one or more parameters to different connections, such that the distribution of use of values of the parameters fits to a predetermined pattern.

An aspect of some embodiments of the invention relates to joint selection of values of a plurality of network parameters for a real-time packet based connection. The joint selection is optionally performed by the same entity and/or based on the same information or information collected by the same entity. In some embodiments of the invention, the selection of the network parameter values is inter-related, i.e., the value selected for a first network parameter is taken into account in setting the value for a second network parameter. For example, after selecting a route, the MTU of the route may be adjusted according to the selection. In an exemplary embodiment of the invention, the decision whether header compression is performed responsive to QoS attributes of the selected route.

An aspect of some embodiments of the invention relates to a method of setting negotiated end-point parameters of a real time packet connection by an intermediary unit of the connection. The method includes intercepting negotiation packets of the connection and changing the parameter values stated in the intercepted packets to desired values. Thus, setting the signaling parameters of connections may be performed based on the needs of the network without requiring compliance of the end-units. Optionally, the connection comprises an over-IP network connection.

There is therefore provided in accordance with an exemplary embodiment of the invention, a method of setting parameters of a real time packet-based connection over a communication network, comprising identifying, by a particular network element, a real-time packet based connection, selecting, by the particular network element, a value for at least one end-point parameter of the identified connection and selecting, by the particular network element, a value for at least one network parameter of the identified connection.

Optionally, the selecting of the values of the end-point parameter and the network parameter is performed during setup of the connection. Optionally, the at least one end-point parameter comprises a negotiated parameter. Alternatively, the at least one end-point parameter comprises at least one non-negotiated parameter. Optionally, the at least one end-point parameter comprises a jitter buffer size, a frame size of transmitted packets on the connection, a codec type and/or a global parameter. Optionally, the at least one network parameter comprises a route to be traversed by the packets of the connection, a header compression method to be applied to the packets of the connection and/or an MTU value of at least one routing unit of the network.

Optionally, the method includes receiving by the particular network element a value of a quality of service QoS attribute of the network, and wherein the selecting of the network

parameter and the end-point parameter is performed responsive to the value of the QoS attribute. Optionally, the value of the QoS attribute is determined before the connection is established. Optionally, the QoS attribute comprises a jitter value of links of the network and/or a delay value of links of the network. Optionally, the QoS attribute comprises an available bandwidth value of links of the network and/or a packet loss value of links of the network.

Optionally, selecting the value for the at least one end-point parameter comprises selecting the value of the end-point parameter responsive to the selected value of the network parameter. Optionally, selecting the value of the network parameter is performed responsive to the selected value of the end-point parameter.

Optionally, selecting the value for the at least one end-point parameter comprises selecting a codec responsive to a delay of a selected route for the connection, such that the total delay of the route in the codec is smaller than a predetermined value. Optionally, selecting of the network parameter value is performed responsive to a type of the connection. Optionally, the method includes transmitting the value of the at least one end-point parameter to an end-point of the connection and transmitting the value of the network parameter to at least one routing unit for implementation thereby. Optionally, the particular network element is not an end-point of the connection. Optionally, the particular network element selects parameter values for a plurality of different connections substantially concurrently.

There is further provided in accordance with an exemplary embodiment of the invention, a method of setting parameters of a real time packet-based connection over a communication network, comprising collecting quality of service attribute values of the network, by one or more network elements, selecting a value for at least one end-point parameter of the connection, responsive to the collected attribute values and selecting a value for at least one network parameter of the connection, responsive to the collected attribute values.

Optionally, the values of the at least one end-point parameter and the at least one network parameter are selected before either of the parameter values is implemented.

There is further provided in accordance with an exemplary embodiment of the invention, a method of setting parameters of a real time packet-based connection over a communication network, comprising determining a value for at least one end-point parameter of the connection and selecting a value for at least one network parameter of the connection, responsive to the determined value of the at least one end-point parameter.

Optionally, the at least one network parameter is selected before the value of the end-point parameter is implemented. Optionally, determining the value of the at least one end-point parameter comprises receiving the value from a unit that determined the value. Alternatively, determining the value of the at least one end-point parameter comprises selecting the value of the at least one end-point parameter.

There is further provided in accordance with an exemplary embodiment of the invention, a method of setting parameters of a real time packet-based connection over a communication network, comprising determining a value for at least one network parameter of the connection and selecting, before implementing the selected value of the network parameter, a value for at least one end-point parameter of the connection, responsive to the determined value of the at least one network parameter.

Optionally, the at least one network parameter comprises a route for the connection.

Optionally, determining the value of the at least one network parameter comprises selecting the value.

There is further provided in accordance with an exemplary embodiment of the invention, an apparatus for determining parameter values of a real time packet-based connection over a communication network, comprising an input interface for receiving information regarding real time connections between end-point units, a processor adapted to select for at least one connection a value for at least one end-point parameter and for at least one network parameter.

Optionally, the input interface is adapted to receive the information by receiving control packets of the connections. Optionally, the processor additionally identifies real-time packet based connections.

There is further provided in accordance with an exemplary embodiment of the invention, a method of setting parameters of a real time packet-based connection over a communication network, comprising collecting quality of service attribute values of the network, by a first network element, selecting a value for at least one non-negotiated end-point parameter of a connection of which a second network element is an end-point, responsive to the collected attribute values and implementing the selected parameter value by the second network element.

Optionally, the non-negotiated end-point parameter comprises a jitter buffer size.

Optionally, selecting the value for the parameter is performed by the first network element. Alternatively, selecting the value for the parameter is performed by the second network element.

There is further provided in accordance with an exemplary embodiment of the invention, a method of setting parameters of a plurality of real time packet-based connections between end-point elements, over a communication network, comprising collecting quality of service attribute values of the network, by a particular network element, selecting, by the particular network element, values for at least one end-point parameter of a plurality of connections over the network and implementing the selected parameter values in respective end-points of the connections.

Optionally, the value of the parameter for at least one connection is selected responsive to the value of the parameter selected for at least one other connection.

There is further provided in accordance with an exemplary embodiment of the invention, a method of setting parameters of a real time packet-based connection over a communication network, comprising receiving, by an intermediary network element, a signaling packet transmitted between end-points of a real time packet based connection, determining a value for at least one end-point parameter of the connection, changing, by the intermediary network element, the value of at least one field of the received signaling packet, responsive to the determined value of the parameter and forwarding, from the intermediary network element, the packet with the changed value.

Optionally, determining the value comprises determining a codec to be used.

Possibly, changing the value of the field comprises changing a field stating codec types supported by an end-point to include only the determined codec.

BRIEF DESCRIPTION OF FIGURES

Particular non-limiting embodiments of the invention will be described with reference to the following description of embodiments in conjunction with the figures. Identical structures, elements or parts which appear in more than one figure are preferably labeled with a same or similar number in all the figures in which they appear, in which:

Fig. 1 is a schematic illustration of a network including parameter setting controllers, in accordance with an exemplary embodiment of the invention;

Fig. 2 is a schematic block diagram of a connection parameter controller, in accordance with an exemplary embodiment of the invention; and

Fig. 3 is a flowchart of acts performed by a connection setup unit upon identifying that a connection is being established, in accordance with an exemplary embodiment of the invention.

DETAILED DESCRIPTION OF EMBODIMENTS

Fig. 1 is a schematic illustration of a network 110 including parameter controllers 120, in accordance with an exemplary embodiment of the invention. Network 110 includes a plurality of routing units 101 (e.g., routers, switches, bridging devices) connected through data links 106. Data links 106 may include substantially any type of transmission link known in the art, including but not limited to satellite, cable, optical fiber, modem lines, point to point lines, dial up links and local area networks. Clients 104 (e.g., computers, telephones, voice mail systems, soft phones, video transmitters and/or receivers) are connected through end-points 102, to the network so as to exchange data packets. The data packets are optionally transferred using the TCP/IP protocol suit or other packet based protocols. At least some of the data passing over the network belong to real time connections carrying telephone, video conferencing, streaming and/or other multimedia data. The multimedia data is transmitted using any suitable protocol known in the art, for example along with signaling in accordance with protocols such as H.323, H.245, session initiation protocol (SIP) and/or session description protocol (SDP). Optionally, gatekeepers 100, as are known in the art, establish multimedia data connections over the network. Alternatively, other call establishment units, such as suitable proxy servers or call agents may be used.

In some embodiments of the invention, parameter controllers 120 are installed in some or all of gatekeepers 100. Alternatively or additionally, parameter controllers 120 are installed on the route between end-points 102 and gatekeepers 100 and/or in parallel to gatekeepers 100, such that controllers 120 intercept connection establishment requests transmitted from end-points 102. Further alternatively or additionally, controllers 120 are located between gatekeepers 100 and a backbone of network 110, intercepting packets transmitted from gatekeepers 100 to remote parties of established connections. Further alternatively or additionally, controllers 120 may be located at other positions along the network. Optionally, routing units 101 are configured to forward connection establishment packets to respective controllers 120. Alternatively or additionally, end-points 102 are configured to transmit signaling packets to gatekeepers 100 and/or to controllers 120.

Controllers 120 are optionally implemented by a suitable software running on a dedicated processor, e.g., a general purpose processor or a signal processing processor.

Alternatively or additionally, controllers 120 are implemented by a software running on a processor of gatekeeper 100 or of other communication elements, such as routers. Further alternatively or additionally, controllers 120 are implemented in hardware, for example by an ASIC, or by a combination of special hardware and software.

Fig. 2 is a schematic block diagram of a controller 120, in accordance with an exemplary embodiment of the invention. Controller 120 optionally includes a network information unit 301, which collects information on the mapping and current QoS condition of the links of the network and a connection setup unit 321, which selects parameter values for connections being established. Network information unit 301 optionally includes a network mapping unit 302 that receives routing information and generates a map of the network. The QoS condition of the network is optionally monitored by a passive monitoring unit 304 and/or by an active monitoring unit 306 so as to generate a network map database 308 that stores information on the QoS condition of the network.

In some embodiments of the invention, network information unit 301 monitors packets of control protocols of the network and according to these packets determines the network information. Optionally, packets of routing protocols, such as routing information protocol (RIP), border gateway protocol (BGP) and open shortest path first (OSPF) are provided to mapping unit 302, which accordingly generates a map of network 110 using methods known in the art. Alternatively or additionally, mapping unit 302 receives link state advertisements (LSA) which are used in determining the network mapping.

Network information unit 301 optionally further receives control packets that describe the QoS condition of the network, such as real-time control protocol (RTCP) packets. The QoS condition packets are optionally provided to passive monitoring unit 304, which generates data of database 308 accordingly, using methods known in the art. In some embodiments of the invention, active monitoring unit 306 transmits test packets (e.g., ping packets) to various network elements and according to the responses to the test packets determines QoS condition information of the network, to be stored in database 308. Optionally, the test packets are transmitted to each of routers 101 in the network. Based on the difference in measurements between adjacent routers 101, the QoS condition of the links of the network are determined, as is known in the art. The QoS condition of the network optionally includes jitter, packet loss, bandwidth and delay.

In some embodiments of the invention, controllers 120 receive periodic updates of the QoS condition of the network from other controllers in the network. Alternatively or

additionally, a central unit (not shown) receives periodic updates from controllers 120 and after processing forwards the information to controllers 120. Further alternatively or additionally, controllers 120 periodically query other controllers for data they need. Alternatively or additionally, a plurality of monitors that monitor the QoS condition of the network and transfer QoS condition information to controllers 120 are distributed throughout the network.

In some embodiments of the invention, network database 308 optionally includes information on network 110 other than the QoS condition, such as the cost of using the links of the network and/or the capabilities of the routers.

Optionally, QoS condition information of the network is collected at a predetermined rate. Alternatively or additionally, the rate at which information is collected depends on the load on the network. Optionally, when the load is high the information is updated at a higher rate in order to better optimize the use of the network when the optimization is important, e.g., when traffic is heavy.

Connection setup unit 321 comprises a signaling analysis unit 310 that keeps track of signaling packets transmitted from end-points 102 passing through controller 120. According to the signaling packets, signaling analysis unit 310 determines when new connections are being established, the types of the connections and end-point capabilities of the end-points 102 establishing the connection. According to the determination of signaling analysis unit 310, a session database 312 is optionally created. A routing unit 314 selects a route for the packets of the connection and/or parameters of the routing units along the route. A configuration unit 316 selects end-point parameters for the connection. In some embodiments of the invention, configuration unit 316 also chooses global parameters of routing units 101, which affect a plurality of connections, such as the MTU of one or more links.

Fig. 3 is a flowchart of acts performed by connection setup unit 321 upon identifying that a connection is being established, in accordance with an exemplary embodiment of the invention. When controller 120 identifies (500) that a connection is being established, controller 120 determines (502) characteristics of the connection, such as the type of the connection and the capabilities of the end-points 102 of the connection. Alternatively or additionally, the characteristics of the end-points are determined before the connection is established.

Responsive to the determined characteristics of the connection, constraints on the QoS of the connection and a function to be optimized in choosing a route for the connection, are

defined (504). A route, which meets the constraints and optimizes the function, is then chosen (506), based on the data in map database 308. The accumulated QoS condition of the selected route is optionally determined (508). According to the determined accumulated QoS condition of the route, end-point parameter values and/or parameter values of routing units 101 are chosen (510). If (512) the resulting QoS of the connection is sufficient, controller 120 instructs (514) end-points 102 and/or routing units 101 on the selected parameter values they are to use. Otherwise, the constraints on the route and/or the optimization function are changed and the choosing (506) of a route is repeated under the new constraints.

Referring in more detail to identifying when a connection is being established (500), in some embodiments of the invention, controller 120 identifies and handles only connections initiated by an end-point 102 serviced by the gatekeeper associated with controller 120. Connections passing through controller 120 initiated remotely are optionally handled, in these embodiments, by the controllers close to the other end-point(s) of the connection.

Alternatively, each controller 120 through which a connection passes handles one of the directions of the connection (upstream or downstream). In an exemplary embodiment of the invention, each controller 120 handles the direction of transmission from the end-point it services. Alternatively, the reception direction is handled by the nearby controller 120.

Further alternatively, each connection is handled by both controllers 120 through which the packets of the connection pass. Optionally, in accordance with this alternative the controllers 120 cooperate and exchange data to determine the acts to be performed. Alternatively, the controllers 120 operate independently. Optionally, controllers 120 apply the same decision making methods and therefore, given the same input data, will make the same decision without cooperation.

Further alternatively, controllers 120 are configured as to which connections they are to handle. Optionally, each controller handles all the connections initiated locally and is configured as to the connections initiated remotely that it is to handle. Further alternatively or additionally, controllers 120 communicate as to which connections they are to handle. For example, controllers 120 may add a signature to signaling packets of connections they handled. Alternatively or additionally, controllers 120 may transmit control packets listing the connections they handle and/or the end-points whose connections they handle.

Controller 120 optionally scans the data passing through it for connection establishment requests, such as the H.323 request-permission-to-place-a-call message.

Alternatively or additionally, end-points 102 are configured to notify controller 120 when they want to establish a call and/or receive a new connection.

Referring in more detail to determining (502) characteristics of the connection, in some embodiments of the invention, the characteristics include the type of the connection (e.g., voice, video). Optionally, controller 120 follows the signaling packets of the setup phase of the connection and determines therefrom the type of the connection.

In some embodiments of the invention, the determined characteristics of the connection include the service level agreement (SLA) pertaining to the connection, as determined for example from records of a service provider of the end-point. Alternatively or additionally, the determined characteristics include the capabilities of the end-points of the connection, for example the codec types supported by the end-points, the maximum number of frames per second supported by the end-points and/or information about support for silence suppression and/or voice activity detection (VAD). The end-point capabilities are optionally determined from H.245 packets, fast connect packets or similar signaling packets of other protocols (e.g., SDP) exchanged on the connection. Alternatively or additionally, the end-point capabilities are determined by querying the end-points by controller 120. In some embodiments of the invention, controller 120 periodically queries the end-points 102 it services, in order to collect data on their capabilities. The data is optionally stored in sessions database 312 for use when a connection is established. Further alternatively or additionally, controller 120 monitors the data transmitted from end-points 102 during previous connections and accordingly determines the capabilities of the end-points. This data is used, for example, when the capability data of the current connection is not available.

Referring in more detail to defining (504) constraints on the QoS of the connection, in some embodiments of the invention, the constraints include a maximal allowable end to end delay of the connection, a maximal allowable jitter, a maximal allowable packet loss rate and/or a maximal allowable quality degradation due to compression. Alternatively or additionally, the constraints for the connection include a minimal bandwidth required for the connection, for example assuming that the strongest compression is used. Optionally, the amount of allowable bandwidth for the connection depends on the expected number of connections expected to pass through the network. In an exemplary embodiment of the invention, the constraint for voice connections includes having a delay beneath a predetermined threshold, e.g., 250 ms. Alternatively or additionally, the constraint for voice connections includes having a jitter below a predetermined threshold, for example 1 ms. In an

exemplary embodiment of the invention, the constraint for video connections includes having a packet loss rate beneath a predetermined threshold, for example, lower than 1%. Alternatively or additionally, video connections have constraints on their delay and/or jitter, as with voice connections.

In some embodiments of the invention, the constraints are defined while taking into account the affects of a predetermined set of values of end-point parameters. The predetermined set of end-point parameter values optionally includes average values of the parameters. For example, a type of codec having an average delay is used in the predetermined set of parameter values used in defining the constraints on the route. Alternatively or additionally, the predetermined set of parameter values includes best values, which are most demanding. As described below, if these parameter values cannot be met, the constraints will be redefined for other parameter values. Further alternatively or additionally, the constraints are defined using worst case parameter values, which are least demanding. If the selected route will allow for better end-point parameter values, such values will optionally be used.

The optimization function optionally includes cost for the service provider or a weighted function of the service provider cost and of a measure of the amount of available bandwidth on the route. Alternatively or additionally, the optimization function depends on the delay and/or jitter of the route. In an exemplary embodiment of the invention, the optimization function is a weighted sum of the factors to be taken into account in the optimization. The weights used are optionally a function of the SLA of the client. For example, the function for a client who pays a low fee gives low weight to delay and high weight to the cost, while the function for a high-fee client gives high weight to the delay.

The choosing (506) of the route is optionally performed using any method known in the art, such as the shortest route computation method, the "branch & bound" method or other methods for finding an optimal route for multiple additive QoS constraints. In some embodiments of the invention, instead of selecting a single route, a plurality of routes with descending preference are chosen together. If, in checking the highest priority chosen route, the route is determined to be unsuitable, a next in priority route is checked and there is no need to repeat the route selection with new constraints.

Optionally, only if none of the selected routes is suitable, the route selection is repeated with new constraints, based on different end-point and/or network parameter values, which have better chances to be operable with the available routes. The new constraints are optionally chosen based on the QoS attributes of the available routes which were not found to

be suitable. Alternatively, the new constraints are based on allowing for a lower quality connection, e.g., with a delay of up to 400 ms instead of 250 ms. Optionally, the optimization function is adjusted to the new constraints. For example, the optimization function when the constraint requires a delay lower than 250 ms optionally does not depend on the delay, while when the constraint requires a delay lower than 400-500 ms, the delay optionally has a substantial weight in the optimization function. Further alternatively, a best route from the selected routes is chosen and if this route is not suitable the connection is temporarily refused.

Alternatively or additionally, the choosing of the route is performed in parallel for a plurality of different sets of constraints, and a first route which meets any of the sets of constraints is used. This alternative is optionally used when the network is relatively loaded and the chances of success in finding a route are relatively low.

Alternatively to choosing a single route, controller 120 may choose routing rules for the connection, such as allowing the packets of the connection to be routed by each of the routers they pass through separately. In some embodiments of the invention, controller 120 may instruct some or all of the network routers as to which algorithm is to be used in determining the next hop of packets belonging to a connection. In an exemplary embodiment of the invention, some or all of the routers support applying a plurality of different routing decision methods, and controller 120 selects one of the methods which is to be applied to the present connection. Alternatively, controller 120 determines whether to impose a specific route or to allow the routers to use a default routing method.

In some embodiments of the invention, a global network parameter set by controllers 120 is the routing method used by one or more of routing units 101 on all the connections they handle.

Referring in more detail to determining (508) the accumulated QoS of the route, in some embodiments of the invention, the accumulated QoS condition of the route is determined by combining the QoS values of the links of the route using methods known in the art. The QoS condition of each link is optionally represented by values of a plurality of different attributes. The accumulated QoS condition of the route optionally includes values for the plurality of attributes, the values of each attribute being combined separately. The QoS condition is optionally represented by one or more of jitter, bandwidth, packet loss and delay.

Referring in more detail to setting (510) end-point parameter values, in some embodiments of the invention, the end-point parameters include the codec type (e.g., for voice, G.729, G.711, G.728, G.723), the frame size (e.g., size of IP packets used), video

resolution, encryption method and/or the jitter buffer size. Optionally, the jitter buffer size is set according to the jitter of the connection. The codec type is optionally selected according to the route delay, according to the time required for compression and decompression, such that the delay is within a desired limit. The codec type is also selected according to the capabilities of the end-points (a codec that is not supported by the end-points is not selected) and optionally the allowed degradation due to compression (e.g., according to the SLA of the client). The frame size is optionally set according to the packet loss, available bandwidth, jitter and/or delay of the route. In some embodiments of the invention, a small frame size is optionally used when there is a relatively high delay, sufficient bandwidth, a high jitter and/or a high packet loss, while a large frame size is used when the route has a low packet loss, limited bandwidth, a low jitter and/or a short delay. The frame size optionally further depends on the application to which the data belongs.

Alternatively or additionally to setting (510) values of end-point parameters, controller 120 sets network parameter values, such as the type of header compression used (if at all) and/or the QoS marking method or policy. The QoS policy of the routers may include, for example, whether the routers relate to QoS signals included in the transmitted packets. In an exemplary embodiment of the invention, when the selected route has a relatively short delay, header compression is used, but header compression is not used when the selected route has a long delay. Alternatively or additionally, header compression is used when required to make a selected route have sufficient bandwidth for the connection.

In some embodiments of the invention, controller 120 also sets values of global network parameters responsive to the QoS condition of the network and/or according to other data on the network operation. Optionally, one of controllers 120 is selected as a master unit that sets global network parameters, so that global parameters are not changed by different controllers without coordination. Alternatively, any time a controller 120 changes a global parameter value, it notifies all the other controllers so that they refrain for a predetermined time (or forever) from changing the value of that parameter.

In some embodiments of the invention, a controller 120 periodically sets the maximal transmission unit (MTU) of the links of the network according to the QoS attributes of the links. Optionally, links which carry a large percentage of voice packets are set to a low MTU and links carrying a large percentage of data transfer packets are set to a high MTU. In some embodiments of the invention, voice packets are identified by the protocol they use. For example, although not always correct, it may be assumed that UDP packets require a low

MTU while TCP packets require a high MTU. Alternatively or additionally, the MTU is lowered on links that have a long delay and a low bandwidth and the MTU is raised on links that have a short delay and a high bandwidth utilization percentage.

Optionally, in some cases when no route meeting the constraints can be found for a connection, a global network parameter is changed in order to make one of the routes meet the constraints. For example, if none of the routes have a short enough delay, the MTU of the route with the shortest delay is optionally reduced in order to improve its delay.

As mentioned above, if (512) no available parameter value is sufficient to bring the QoS condition of the route into a range of desired limitations (not necessarily the same as the constraints used in the route selection), the constraints for route selection are changed (516) and the choosing (506) of a route is repeated. Optionally, the route is checked as to whether it can carry the bandwidth required for the packets of the connection (according to the codec and/or frame size used). Alternatively or additionally, the route is checked as to whether its jitter is reasonable. Further alternatively or additionally, the route is checked as to whether its delay is suitable.

Alternatively to repeating the selection of a route if one of the conditions is not met, one or more parameters are adjusted, if possible, in order to meet the requirements. For example, when bandwidth is lacking, routing units 101 may be instructed to perform header compression.

Referring in more detail to instructing (514) end-points 102 and/or routing units 101 on their parameter values, in some embodiments of the invention, controller 120 transmits configuration instructions to end-points 102 and/or routing units 101, using predetermined protocol rules as are known in the art. Alternatively or additionally, controller 120 intercepts signaling packets exchanged between the end-points 102 of the connection and changes the values of one or more parameters in the signaling packets. In some embodiments of the invention, the signaling packet is changed to indicate support of only a single parameter value, i.e., the value determined to be used by controller 120. For example, the codec-type field of the signaling packet, indicating the codec types supported by the end-point transmitting the packet, is optionally changed to indicate only the codec type selected for the connection by controller 120. Alternatively or additionally, the signaling packet is changed to indicate a different order of codec types, according to the order of priority selected by controller 120.

In an exemplary scenario, the signaling includes transmission of three packets between the end-points of the connection. The three packets include a first packet transmitting possible

values from the initiator end-point of the connection, a second packet responding to these values by the remote end-point of the connection with possible values of the remote end-point and a final selection packet from the initiator. In some embodiments of the invention, controller 120 changes the first packet indicating the capabilities of the initiator end-point, so that the end-point will choose a desired parameter value. Alternatively or additionally, controller 120 changes the second packet with the response from the remote end-point.

In some embodiments of the invention, the route is imposed by stating in each of the packets of the connection the route it is to traverse, for example using the IP source routing option. Optionally, the route is inserted to the packet by the end-point of the connection. Alternatively, a gatekeeper or other unit at the entrance to network 100 adds the route information to the packets of the connection. Alternatively, the route is imposed using a reservation protocol, such as resource reservation protocol (RSVP), multiple protocol label switching (MPLS) or constraint based routing using label distribution protocol (CR-LDP). Alternatively or additionally, other methods are used to impose the route, such as using a redirect method. In some embodiments of the invention, the protocol is implemented by the end-points under the instructions of controller 120. Alternatively, the protocol is implemented by controller 120 without the end-points being aware of the setting of the route.

The above described method of Fig. 3 is presented by way of example, and many other methods of setting the parameter values of connections may be used. For example, rather than first selecting a route and then selecting end-point parameter values accordingly, the end-point parameters may be selected first and the route selected accordingly.

Following is an exemplary implementation of a parameter selection method in accordance with an embodiment of the invention. A first node A desires to establish a voice over IP (VoIP) connection with a second node B. It is assumed that the network has two routes between node A and node B. A first route has two hops and has an accumulative QoS of a delay of 140 ms, a jitter of 10 ms, a packet loss of 1% and an available bit rate of 50 kbps. A second route has three hops and has an accumulative QoS of a delay of 128 ms, a jitter of 20 ms, a packet loss of 2% and an available bit rate of 80 kbps.

It is assumed in this example that the end-to-end requirements of the VoIP connections are a delay smaller than 150ms, a jitter of less than 20 ms and a packet loss of less than 3%. Node A advertises support of G.711, G.728 and G.729 codecs and node B advertises support of G.711, G.723 and G.729 codecs.

Controller 120 optionally first tries using the G.729 codec, because it requires less bandwidth than the G.711 codec. The G.729 codec requires 30 kbps, uses 50 packets per second (pps) and adds a delay of 25ms for compression and decompression. Using the G.729 codec leaves for the route 150ms-25ms=125ms, which is not available in either the first or second route. Controller 120 applies a route search method, such as "branch & bound", under the constraints that the route has at least 30 kbps of available bandwidth and has a delay smaller than 125 ms. According to the above exemplary numbers no route will be found and therefore controller 120 will try the G.711 codec, which adds only a delay of 0.2 ms but requires about 80kbps of available bandwidth. The first route will not be selected due to a lack of available bandwidth, so the second route will be used. The jitter buffer will be configured accordingly to a size of 20ms.

In some embodiments of the invention, the above described method of selecting parameters is performed for all connections passing on the network, regardless of the current state of the network. Thus, even if the current load does not require the optimization method of the present invention, in case a sudden flux of connection establishment requests occurs, the network utilization will be optimal. Alternatively, when the network is relatively not congested, controllers 120 do not interfere with the operation of the network and allow regular route selection and signaling. Thus, processing resources are not spent by controllers 120 when the chances that the optimization will be of importance is relatively slight. Alternatively or additionally, the method of Fig. 3 is implemented only at specific times of the day, e.g., during work hours, or on specific dates, responsive to a command from an operator or according to any other parameter external to the network. Further alternatively or additionally, when the load is relatively low or at specific times, only some of controllers 120 implement the method of Fig. 3, or the method is implemented on only some of the connections of one or more of the controllers. For example, the method may be implemented on a predetermined percentage of the connections passing through the controller or whenever the controller has available processing resources. Alternatively or additionally, the method is applied on specific types of connections and/or on connections having a specific QoS or SLA.

In some embodiments of the invention, the same parameter selection method is applied by all controllers 120 of the network on all the connections they handle. Alternatively, different parameter selection methods are applied by different controllers 120 of the same network and/or different methods are applied to different connections by the same controller, for example according to the SLA of the connections.

In some embodiments of the invention, in addition to setting the network and end-point parameters of network 110, controllers 120 provide reports on the operation of the network. The reports optionally include details on the network usage, such as real-time performance reports, service level management reports and application reports. Alternatively or additionally, the provided reports include trend analysis for the entire network or portions thereof. Further alternatively or additionally, reports are generated about the throughput and congestion performance of the network and/or QoS condition reports.

The selection of the parameter values at the time the connection is established allows simple implementation of the values. It is noted, however, that in some cases, for example when the QoS condition of the network changes substantially, the end-point parameters and/or the network parameters may be changed during transfer of data on a connection. Optionally, controllers 120 monitor the QoS conditions of the connections they handle. If the QoS conditions change substantially, for example the delay and/or jitter increases, so as to lower the QoS of the connection beneath a minimal allowed threshold, controller 120 changes the value of one or more parameters. Alternatively or additionally, for example if change in the value of one parameter is not sufficient to avoid the degradation in QoS, the procedure of Fig. 3 is repeated. During the parameter value determination, the connection optionally continues to operate using the old parameter values. After the parameter value selection is completed, the parameter values are changed, optionally in parallel.

Although in the above description the decision on the parameter values and connection routes are performed by controller 120, other units may select the parameter values instead of the controller. In some embodiments of the invention, a central unit receives information from controllers 120 and performs the parameter selection. Thus, it is not necessary that each controller 120 manage a map of the network. Optionally, in these embodiments, network information unit 301 includes only network map database 308 and a port for receiving data from the central unit. In some embodiments of the invention, the routing parameters and end-point parameters are selected by separate units which cooperate. For example, controller 120 may select the route and then notify the end-point 102 on the QoS condition of the route, so that the end-point will choose the end-point parameters according to the selected route. In an alternative embodiment, controller 120 selects the end-point parameter values and a central unit chooses a route accordingly.

It will be appreciated that the above described methods may be varied in many ways, including, changing the order of steps, and the exact implementation used. The methods of the

present invention may be performed in various protocol layers and may be performed for a single transmission system in a plurality of communication protocol layers. It should also be appreciated that the above described description of methods and apparatus are to be interpreted as including apparatus for carrying out the methods and methods of using the apparatus.

The present invention has been described using non-limiting detailed descriptions of embodiments thereof that are provided by way of example and are not intended to limit the scope of the invention. For example, the tasks related above to end-points may be performed by conversion units which establish UDP connections, such as proxies and load balancers. It should be understood that features and/or steps described with respect to one embodiment may be used with other embodiments and that not all embodiments of the invention have all of the features and/or steps shown in a particular figure or described with respect to one of the embodiments. Variations of embodiments described will occur to persons of the art.

It is noted that some of the above described embodiments may describe the best mode contemplated by the inventors and therefore may include structure, acts or details of structures and acts that may not be essential to the invention and which are described as examples. Structure and acts described herein are replaceable by equivalents which perform the same function, even if the structure or acts are different, as known in the art. Therefore, the scope of the invention is limited only by the elements and limitations as used in the claims. When used in the following claims, the terms "comprise", "include", "have" and their conjugates mean "including but not limited to". The term “value” refers herein to one of a plurality of possibilities to which a parameter may be set and is not limited to numerical possibilities.